

Statistica Sinica Preprint No: SS-2021-0138

| | |
|--|---|
| Title | Functional Response Quantile Regression Model |
| Manuscript ID | SS-2021-0138 |
| URL | http://www.stat.sinica.edu.tw/statistica/ |
| DOI | 10.5705/ss.202021.0138 |
| Complete List of Authors | Xingcai Zhou, Dehan Kong, A. B. Kashlak, Linglong Kong, R. Karunamuni and Hongtu Zhu |
| Corresponding Author | Linglong Kong |
| E-mail | lkong@ualberta.ca |
| Notice: Accepted version subject to English editing. | |

Functional Response Quantile Regression Model

Xingcai Zhou¹, Dehan Kong², A. B. Kashlak³, Linglong Kong³,
R. Karunamuni³, and Hongtu Zhu⁴

¹*Institute of Statistics and Data Science, Nanjing Audit University,
Nanjing 211815, China*

²*Department of Statistical Sciences, University of Toronto,
Toronto, Ontario M5S 3G3, Canada*

³*Department of Mathematical and Statistical Sciences, University of Alberta,
Edmonton, Alberta T6G 2G1, Canada*

⁴*Department of Biostatistics, University of North Carolina at Chapel Hill,
Chapel Hill, North Carolina 27516, U.S.A.*

Abstract: In this paper, we propose a new functional response quantile regression model. A data driven estimation procedure is developed to estimate the quantile regression processes based on local linear approximation. Theoretically, we obtain the global uniform Bahadur representation of the estimator with respect to the time/location and the quantile level, and show that the estimator converges weakly to a two-parameter continuous Gaussian process. We then derive the asymptotic bias and mean integrated squared error of smoothed individual functions and their uniform convergence rates under given quantile levels. Based

on the theoretical results, we introduce a global test for the coefficient functions and discuss how to construct simultaneous confidence bands. We evaluate our method through simulations and two applications from diffusion tensor imaging data and ADHD-200 functional magnetic resonance imaging data.

Key words and phrases: Functional data, Global test statistic, Simultaneous confidence band, Weak convergence.

1. Introduction

Functional data analysis deals with data that are in the form of functions, images and shapes, or more general objects [29]. To model the functional data, functional regression models are widely used, including functional linear regression [24, 25, 33] and functional response regression model [25]. Classical functional linear regression describes the relationship between a scalar response and a functional predictor; while the functional response regression characterizes the relation between a functional response and scalar predictors.

The functional response regression model is defined as independent realizations of an underlying stochastic process

$$y_i(s) = x_i^T \beta(s) + \eta_i(s) \quad (i = 1, \dots, n) \quad (1.1)$$

where $y_i(s)$ denotes a functional response for the i th subject, x_i is its as-

sociated p -dimensional covariates of interest, $\beta(s) = \{\beta_1(s), \dots, \beta_p(s)\}^T$ is a $p \times 1$ unknown smooth function of $s \in \mathcal{I}$, $\eta_i(s)$ includes an individual variation, which is used to characterize the within-curve dependence. This functional response model can be used in neuroimaging applications, where one is interested in using clinical, genetic and neuropsychological assessment to predict the time/location varying brain signals. This model is also closely connected to various varying-coefficient models [11, 27, 35, 34, 39]. Typically brain signals are distorted by artifacts and noise, where outliers frequently occur. Studies have found that outliers may affect the statistical results and conclusions significantly [18, 9]. Quantile regression [15] has emerged as an important statistical methodology, which can provide robust statistical results when outliers exist. It allows scientists to make statistical inference on the entire conditional distribution by estimating a collection of conditional quantiles, and does not require specifying any error distribution. Therefore, it has been used widely in various disciplines such as biology, medicine, finance and economics, especially to model complex data such as longitudinal and functional data. A comprehensive survey of quantile regression can be found in [14].

To this end, we propose a *functional response quantile regression model*

(FRQR):

$$y_i(s) = x_i^T \beta(s, \tau) + \eta_i(s, \tau), \quad (i = 1, \dots, n) \quad (1.2)$$

where $\tau \in (0, 1)$ is a given quantile level, and $\eta_i(s, \tau)$ is a stochastic process with covariance function $\gamma_\eta(s, t)$ at each τ . Without loss of generality, we assume that the τ th quantile of $\eta_i(s, \tau)$ equals zero, that is, $F_{\eta_i(s, \tau)}^{-1}(\tau) = 0$, where $F_{\eta_i(s, \tau)}^{-1}$ is the quantile function of $F_{\eta_i(s, \tau)}$, and $F_{\eta(s, \tau)}$ denotes the distribution function of $\eta_i(s, \tau)$ for any s and a given τ . Thus the τ th conditional quantile of $y_i(s)$ given x_i can be written as $Q_{y_i(s)}(\tau | x_i, s) = x_i^T \beta(s, \tau)$, which is characterized by the only parameter $\beta(s, \tau)$. In fact, if $F_{\eta_i(s, \tau)}^{-1}(\tau) \neq 0$, then $(x_i, y_i(s))$ obeys the following conditional quantile restriction $Q_{y_i(s)}(\tau | x_i, s) = F_{\eta_i(s, \tau)}^{-1}(\tau) + x_i^T \beta(s, \tau)$. In quantile regression, each conditional quantile is very important for characterizing the distribution of response. For the fixed s , the model (1.2) becomes the usual quantile regression [14]. Therefore, it is identifiable. In addition, the covariance structure of $y_i(\cdot)$ at the τ th quantile level is characterized by $\eta_i(\cdot, \tau)$ ' covariance function. We introduce an efficient estimation procedure for $\beta(s, \tau)$ based on local linear approximation, derive the global uniform Bahadur representation of $\hat{\beta}(s, \tau)$ and establish the weak convergence results of $\hat{\beta}(s, \tau)$. Based on the results, we propose statistical inference procedures including a global test statistic and simultaneous confidence bands (SCBs). The pro-

cedure has taken into account the within-curve dependence structure and the estimation errors of the underlying unknown data distributions.

Recently, the FRQR model (1.2) has received increasing attention. Liu et al. [19] performed quantile regression on functional response in a Bayesian framework, and applied it to the mass spectrometry proteomics data; Yang et al. [30] proposed a Bayesian quantile functional regression, via the quantile functions as functional data, to model the entire marginal distribution of pixel intensities of the tumor images; Zhang et al. [36] developed a novel spatial FRQR model to characterize the conditional distribution of the image response on the whole spatial domain. Liu et al. [20] provided a comprehensive understanding of how scalar predictors influence the conditional distribution of functional response by using quantile regression, and also performed statistical inference via asymptotic simultaneous confidence bands (SCBs). Our work differs substantially from [20] in the following aspects. First, Liu et al. [20] performed quantile regression separately at each sampling location to obtain the pointwise estimator of the coefficient functions, and then constructed SCBs based on linear interpolation (LI). In contrast, we apply kernel smoothing (KS) to estimate the regression coefficient functions, directly perform the global test for linear hypotheses of the coefficient functions and construct the SCB for each coef-

ficient function. The estimation method of [20] is only based on two points on both sides of the point of interest, which does not utilize full information from its neighbors. Thus, the SCB obtained by the LI method may need to be corrected in practice. Second, Liu et al. [20] did not consider the global hypothesis test. Third, the theoretical development in our work is different from the one in [20]. In particular, Liu et al. [20] gave a uniform Bahadur representation of pointwise estimator on the discrete sampling grid, and derived a strong Gaussian approximation of LI estimator over the location $s \in \mathcal{I}$ for the fixed quantile level $\tau \in (0, 1)$. In contrast, we obtain the global uniform Bahadur representation of the KS estimator, and establish a strong two-parameter Gaussian approximation of the estimator with respect to $s \in \mathcal{I}$ and $\tau \in (0, 1)$. Our theoretical development is more challenging than Liu et al. [20]'s.

We have made the following contributions. First, as far as we are aware, compared to existing literatures on functional quantile regression [13, 4], this is the first time that functional response is considered in quantile regression literature. Second, we have developed novel theories, such as the global uniform Bahadur representation and weak convergence of the estimator of varying-coefficient function, using advanced empirical process methods. To validate our inference procedure, we have also derived the

null distribution of the global test statistic and introduced a wild bootstrap testing procedure with theoretical guarantees. Third, our approach reveals interesting findings in two applications from diffusion tensor imaging data and ADHD-200 functional magnetic resonance imaging data.

2. Estimation procedure

2.1 Assumptions

Throughout this paper we assume $s \in [0, 1]$, but our results can be easily extended to the unit square, cube, or higher dimensions. We assume that $\eta_i(s, \tau)$ are independent identical copies of the stochastic process $\eta(s, \tau)$ with covariance function $\gamma_\eta(s, t)$ at each τ and whose τ th quantile equals zero. For imaging data, it is typical that functional responses $y_i(s)$ are measured at the same location for all subjects. Therefore, $y_i(s)$ is measured at the same m location points $0 = s_1 \leq s_2 \leq \dots \leq s_m = 1$ for all $i = 1, \dots, n$. We also assume $\beta(\cdot, \tau)$ is twice continuously differentiable.

2.2 Quantile regression estimation

In model (1.2), the conditional quantile function of $y_i(s)$ given x_i can be expressed as $Q_{y_i(s)}(\tau \mid x_i, s) = x_i^T \beta(s, \tau)$, $i = 1, \dots, n$ for each $s \in \mathcal{I}$ and any $\tau \in (0, 1)$. The main parameter of interest is $\beta(s, \tau)$, where we employ

2.3 Data-driven estimation of density and distribution functions

local polynomial regression [6] to estimate it. In particular, $\beta(s_j, \tau)$ can be locally approximated by a linear function as $\beta(s_j, \tau) \approx \beta(s, \tau) + \dot{\beta}(s, \tau)(s_j - s) \equiv b_1 + b_2(s_j - s)$, where $\dot{\beta}(s, \tau) = \{d\beta_1(s, \tau)/ds, \dots, d\beta_p(s, \tau)/ds\}^T$. Let $z_j(s) = z_{h_1j}(s) = \{1, (s_j - s)/h_1\}^T$, $z_{x,ij}(s) = \{x_i^T, h_1^{-1}(s_j - s)x_i^T\}^T = z_j(s) \otimes x_i$, where \otimes is the Kronecker product, $b = (b_1^T, h_1 b_2^T)^T$, $K(s)$ is a kernel function, $K_{h_1j}(s) = h_1^{-1}K_j(s)$ with $K_j(s) = K\{(s_j - s)/h_1\}$, and h_1 is a bandwidth. Then, the estimator $\hat{\beta}(s, \tau)$ can be obtained by minimizing the following locally weighted quantile regression loss function

$$\sum_{i=1}^n \sum_{j=1}^m \rho_\tau \{y_i(s_j) - z_{x,ij}(s)b\} K_j(s) \quad (2.3)$$

where $\rho_\tau(u) = u\{\tau - I(u < 0)\}$ is a check function [15]. One can see

$$\hat{\beta}(s, \tau) = \{(1, 0) \otimes I_p\} \hat{b}.$$

For choice of h_1 , we introduce an automatic selection procedure in Section 2.4.

2.3 Data-driven estimation of density and distribution functions

We need to estimate the unknown error densities of $\eta(s)$, and the unknown bivariate cumulative distribution of individual effects $\eta(s)$ and $\eta(t)$, for each $s, t \in [0, 1]$. We achieve such estimation by the residual-based empirical distributions. It is useful to estimate the distribution of the error

2.3 Data-driven estimation of density and distribution functions

by the empirical distribution of residuals, and to develop tests for model assumptions by applying this estimated error distribution [1, 22, 23].

Define the bivariate cumulative distribution of individual effects $\eta(s)$ and $\eta(t)$ as $F_\eta(a_1, a_2, s, t; \tau, \iota) = \text{pr}\{\eta(s, \tau) < a_1, \eta(t, \iota) < a_2\}$, and denote $f_\eta(a_1, a_2, s, t; \tau, \iota) = \frac{\partial^2 F_\eta(a_1, a_2, s, t; \tau, \iota)}{\partial a_1 \partial a_2}$. Exactly, we need to estimate $f_\eta(0, s; \tau)$ and $\hat{F}_\eta(0, 0, s, t; \tau, \iota)$ for $s \neq t$ for statistical inference.

First, we can estimate $f_\eta(a, s; \tau)$ by kernel methods such as

$$\hat{f}_\eta(a, s; \tau) = n^{-1} \sum_{i=1}^n K_{h_2}\{a - \hat{\eta}_i(s, \tau)\}$$

with the residuals $\hat{\eta}_i(s, \tau) = y_i(s) - x_i^\top \hat{\beta}(s, \tau)$. Thus, we get $\hat{f}_\eta(0, s; \tau)$.

Second, we can estimate the multivariate cumulative distribution function $F_\eta(a_1, a_2, s, t; \tau, \iota)$ via the kernel estimation such as

$$\hat{F}_\eta(a_1, a_2, s, t; \tau, \iota) = \int_{-\infty}^{a_1} \int_{-\infty}^{a_2} \hat{f}_\eta(b_1, b_2, s, t; \tau, \iota) db_1 db_2$$

for $s \neq t$, with

$$\hat{f}_\eta(b_1, b_2, s, t; \tau, \iota) = n^{-1} \sum_{i=1}^n K_{h_3}\{b_1 - \hat{\eta}_i(s, \tau)\} K_{h_4}\{b_2 - \hat{\eta}_i(t, \iota)\}.$$

So, we obtain $\hat{F}_\eta(0, 0, s, t; \tau, \iota)$.

For statistical inference, the unknown error density needs to be estimated. Popular ways include the difference quotient method [12, 28] and the Nadaraya-Watson type or local linear kernel method [7]. In this arti-

2.4 Automatic bandwidth selection

cle, we adopt the Nadaraya-Watson type estimators for the error processes' densities $f_\eta(\cdot, s; \tau)$ and $f_\eta(\cdot, \cdot, s, t; \tau, \iota)$.

2.4 Automatic bandwidth selection

Bandwidth selection is critical in local smoothing. To implement our estimation method, we need to choose appropriate bandwidths h_l ($l = 1, \dots, 4$), for obtaining adequate estimators $\hat{\beta}(s)$, $\hat{f}_\eta(a, s; \tau)$ and $\hat{f}_\eta(b_1, b_2, s, t; \tau, \iota)$. Asymptotically, we require h_l ($l = 1, \dots, 4$) satisfying the conditions (C7)-(C9) in Section 4. It is difficult to use them in practice. A simple and efficient way to obtain these bandwidths is to use cross-validation (CV) based on data. For the Nadaraya-Watson estimator of random error density, bandwidth selection often minimizes the least integrated squared cross-validation (CV) score [8, 10]. In this article, we adopt minimizing the sum of the integrated squared CV score to choose bandwidth for the Nadaraya-Watson estimator of the stochastic process density function.

For h_1 , we minimize the CV score,

$$\text{CV}(h_1) = \sum_{i=1}^n \sum_{j=1}^m \rho_\tau \{y_i(s_j) - x_i^\top \hat{\beta}(s_j, \tau, h_1)^{(-i)}\},$$

where $\hat{\beta}(s_j, \tau, h_1)^{(-i)}$ is the locally weighted quantile estimator of $\beta(s, \tau)$ with the bandwidth h_1 based on deleting the i th subject from the data.

Following the heuristic suggestions of [26], h_l ($l = 2, 3, 4$) can be chosen

2.4 Automatic bandwidth selection

by cross validation. [8] and [10] proposed cross validation for nonparametric conditional density estimators. Here, we adopt their method to our estimations of density functions. For an estimator $\hat{f}_\eta(a, s_l; \tau)$ of $f_\eta(a, s_l; \tau)$, $s_l \in \mathcal{S}$ and $\tau \in \mathcal{I}$, then define the integrated squared error

$$\begin{aligned} I(s_l, \tau) &= \int \left\{ \hat{f}_\eta(a, s_l; \tau) - f_\eta(a, s_l; \tau) \right\}^2 da \\ &= \int \hat{f}_\eta^2(a, s_l; \tau) da - 2 \int \hat{f}_\eta(a, s_l; \tau) f_\eta(a, s_l; \tau) da + \int f_\eta^2(a, s_l; \tau) da \\ &= I_1(s_l, \tau) - I_2(s_l, \tau) + I_3(s_l, \tau). \end{aligned}$$

Note that $I_3(s_l, \tau)$ does not depend on the bandwidth h_2 and $f_\eta(a, s_l; \tau)$ in $I_2(s_l, \tau)$ is unknown. So, the CV estimator of

$$\hat{I}_2(s_l, \tau) = \frac{2}{n} \sum_{i=1}^n \hat{f}_\eta^{(-i)} \{ \hat{\eta}_i(s_l, \tau), s_l; \tau \} = \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{j=1, j \neq i}^n K_{h_2} \{ \hat{\eta}_i(s_l, \tau) - \hat{\eta}_j(s_l, \tau) \},$$

where $\hat{f}_\eta^{(-i)} \{ \hat{\eta}_i(s_l, \tau), s_l; \tau \}$ denotes the estimator $f_\eta(a, s_l; \tau)$ based on data with the observations of the i th subject left out at s_l .

For h_3 and h_4 , minimize the sum of the integrated squared CV score:

$$\begin{aligned} &\text{SCV}(h_3, h_4) \\ &= \sum_{k, l=1}^m \left[\iint \hat{f}_\eta^2(b_1, b_2, s_k, s_l; \tau, \iota) db_1 db_2 - \frac{2}{n} \sum_{i=1}^n \hat{f}_\eta^{(-i)} \{ \hat{\eta}_i(s_k, \tau), \hat{\eta}_i(s_l, \iota), s_k, s_l; \tau, \iota \} \right] \end{aligned}$$

where $\hat{f}_\eta^{(-i)} \{ \hat{\eta}_i(s_k, \tau), \hat{\eta}_i(s_l, \iota), s_k, s_l; \tau, \iota \} = (n-1)^{-1} \sum_{j=1, j \neq i}^n K_{h_3} \{ \hat{\eta}_i(s_k, \tau) - \hat{\eta}_j(s_k, \iota) \} K_{h_4} \{ \hat{\eta}_i(s_l, \tau) - \hat{\eta}_j(s_l, \iota) \}$. In practice, we assume that $h_3 = h_4$ to save computational time.

The above each bandwidth choice is common in kernel estimation. Based on simple grid search, bandwidths h_1 , h_2 and h_3 are selected adaptively one by one via the above CV methods. In simulation study, we found that the bandwidth h_1 is slightly sensitive to the estimators, and h_l ($l = 2, 3$) are not sensitive to statistical inference. On the whole, the results via our CV selection are satisfactory from our simulation study and real data analysis.

It is better to highlight that the theory developed in the article does not support the procedure with tuning parameters selected adaptively

3. Inference procedure

3.1 Global test statistic

In this subsection, we propose a global test statistic for the general hypothesis testing problem

$$H_0 : C\beta(s, \tau) = c(s, \tau) \text{ for all } s \text{ and } \tau \text{ versus } H_1 : C\beta(s, \tau) \neq c(s, \tau), \quad (3.4)$$

where C is a given $r \times p$ full rank matrix, and $c(s, \tau) = \{c_1(s, \tau), \dots, c_r(s, \tau)\}^T$ is a given vector of functions. Denote $e_{l,p}$ to be the p -dimensional unit vector whose l th component is 1 and others are 0. If we take $C = e_{l,p}^T = (0, \dots, 0, 1, 0, \dots, 0)$ and $c(s, \tau) = 0$, it tests the significance of the l th co-

3.1 Global test statistic

variate effect on the functional responses; if we take $C = (e_{k,p} - e_{l,p})^T, k \neq l$, and $c(s, \tau) = 0$, it tests for an equality $\beta_k(s, \tau) = \beta_l(s, \tau)$.

From Theorem 1, $n^{1/2}[\hat{\beta}(s, \tau) - \beta(s, \tau) - \text{Bias}\{\hat{\beta}(s, \tau)\}]$ converges weakly to a centered Gaussian process $\mathbb{G}(s, \tau)$ with covariance $\Sigma(s, t; \tau, \iota) = E\{\mathbb{G}(s, \tau)\mathbb{G}^T(t, \iota)\}$, indexed by $s \in [0, 1]$ and $\tau \in (0, 1)$. Let the residual process $d(s, \tau) = n^{1/2}[C[\hat{\beta}(s, \tau) - \text{Bias}\{\hat{\beta}(s, \tau)\}] - c(s, \tau)]$, and the normalized version $d_c(s, \tau) = \{C\hat{\Sigma}(s, s; \tau, \tau)C^T\}^{-1/2}d(s, \tau)$, where $\hat{\Sigma}(\cdot, \cdot; \cdot, \cdot)$ is an estimator of $\Sigma(\cdot, \cdot; \cdot, \cdot)$, $\hat{\Sigma}(s, t; \tau, \iota) = (1, 0)\hat{A}^{-1}(s, \tau)\hat{G}(s, t; \tau, \iota)\hat{A}^{-1}(t, \iota)(1, 0)^T \hat{\Omega}_x^{-1}$, $\hat{\Omega}_x = \frac{1}{n} \sum_{i=1}^n x_i x_i^T$, $\hat{A}(s, \tau) = \frac{1}{m} \sum_{j=1}^m K_{h_{1j}}(s) \hat{f}_\eta(0, s_j; \tau) z_j(s) z_j^T(s)$, and 2×2 matrix $\hat{G}(s, t; \tau, \iota)$ with entries

$$\hat{G}_{kl}(s, t; \tau, \iota) = \begin{cases} \hat{\mu}_k(K; s, h_1) \hat{\mu}_l(K; t, h_1) (\tau \wedge \iota - \tau \iota), & \text{if } s = t, \\ \hat{\mu}_k(K; s, h_1) \hat{\mu}_l(K; t, h_1) \{\hat{F}_\eta(0, 0, s, t; \tau, \iota) - \tau \iota\}, & \text{if } s \neq t, \\ (k, l = 0, 1), \end{cases}$$

$$\hat{\mu}_k(K; s, h_1) = m^{-1} \sum_{j=1}^m K_{h_{1j}}(s) \left(\frac{s_j - s}{h_1} \right)^k \quad (k = 0, 1, 2).$$

From weak convergence and the uniform consistency of $\hat{\Sigma}(s, t; \tau, \iota)$ in $[0, 1]^2 \times (0, 1)^2$, we can see $d_c(s, \tau)$ converges weakly to $\mathbb{G}_c(s, \tau) = \{C\Sigma(s, s; \tau, \tau)C^T\}^{-1/2} C\mathbb{G}(s, \tau)$ by Slutsky's theorem [17], where $\mathbb{G}_c(s, \tau)$ is a centered Gaussian process.

The global test statistic for the linear hypothesis H_0 is defined as follows:

$$T_n = \int_0^1 \int_0^1 d_c^T(s, \tau) d_c(s, \tau) ds d\tau. \text{ Let } \tilde{T}_n = \int_0^1 \int_0^1 \mathbb{G}_c^T(s, \tau) \mathbb{G}_c(s, \tau) ds d\tau. \text{ We}$$

3.1 Global test statistic

can show that T_n converges weakly to \tilde{T}_n in Theorem 3.

Since $\text{Bias}\{\hat{\beta}(s, \tau)\}$ is unknown, we can estimate it by

$$\begin{aligned} & \widehat{\text{Bias}}\{\hat{\beta}(s, \tau)\} \\ &= - (I_p, 0) \left\{ \hat{A}^{-1}(s, \tau) \otimes \hat{\Omega}_x^{-1} \right\} (nm)^{-1} \sum_{i=1}^n \sum_{j=1}^m K_{h_{1j}}(s) z_{x,ij}(s) \hat{e}_{ij}(s, \tau), \end{aligned} \quad (3.5)$$

where $\hat{e}_{ij}(s, \tau) = -2^{-1} x_i^T \hat{\hat{\beta}}(s, \tau) (s_j - s)^2 - 6^{-1} x_i^T \hat{\hat{\beta}}(s, \tau) (s_j - s)^3$ in which $\hat{\hat{\beta}}(s, \tau)$ and $\hat{\hat{\beta}}(s, \tau)$ are obtained by using local cubic fit with a selected bandwidth.

In practice, it is difficult to directly obtain the percentiles of T_n even if its null distribution is known. Therefore, we use a wild bootstrap method to approximate critical values of T_n . In particular, we fit model (1.2) under H_0 to obtain $\hat{\beta}^*(s_j, \tau)$ and $\hat{\eta}_{i,0}^*(s_j, \tau)$ for $i = 1, \dots, n$, $j = 1, \dots, m$. Then, we generate a random sample $\tilde{\zeta}_i(s_j)$ from a standard normal distribution for $i = 1, \dots, n$, $j = 1, \dots, m$, and compute $\tilde{y}_i(s_j) = x_i^T \hat{\beta}^*(s_j, \tau) + \tilde{\zeta}_i(s_j) \hat{\eta}_{i,0}^*(s_j, \tau)$. From the $\tilde{y}_i(s_j)$, we reobtain $\tilde{\beta}(s, \tau)$, $\text{Bias}\{\tilde{\beta}(s, \tau)\}$ and $d^{(B)}(s, \tau) = n^{1/2}[C[\tilde{\beta}(s, \tau) - \text{Bias}\{\tilde{\beta}(s, \tau)\}] - c(s, \tau)]$ by using the locally weighted quantile regression loss function (2.3). As we show in Theorem 1, $\text{Bias}\{\tilde{\beta}(s, \tau)\}$ is asymptotically negligible, we drop the term $\text{Bias}\{\tilde{\beta}(s, \tau)\}$ for computational efficiency. Specifically, we calculate $\tilde{T}_n = \int_0^1 \int_0^1 d_c(s, \tau)^T$

3.2 Simultaneous confidence bands

$d_c(s, \tau) ds d\tau$, where $d_c(s, \tau) = \{C\tilde{\Sigma}(s, s; \tau, \tau)C^T\}^{-1/2}d(s, \tau)$. Repeat M times to have $\{\tilde{T}_{n,B}, B = 1, \dots, M\}$, and compute the p -value as $M^{-1} \sum_{B=1}^M I(\tilde{T}_{n,B} \geq T_n)$. If the p -value is smaller than a given significance level α , say 0.05, then one rejects H_0 .

Although we theoretically obtain a global test statistic T_n over $s \in [0, 1]$ and $\tau \in (0, 1)$, in practice, the required double integration for T_n is computationally expensive. In practice, we are often more interested in the function of s for a given τ . Therefore, we may consider the global test statistic $T_n^* = \int_0^1 d_c^T(s, \tau) d_c(s, \tau) ds$ of $\{\hat{\beta}(s, \tau), s \in [0, 1]\}$ for a given $\tau \in (0, 1)$. The null distribution of T_n^* is given in Theorem 3 (b), and similar wild bootstrap procedure can be implemented to improve computational efficiency.

3.2 Simultaneous confidence bands

For a given quantile τ and a preassigned significance level α , we construct simultaneous confidence bands of $\beta_l(s, \tau)$ for each τ , i.e. we find $\hat{\beta}_l^{L,\alpha}(s, \tau)$ and $\hat{\beta}_l^{U,\alpha}(s, \tau)$ such that

$$\text{pr}\{\hat{\beta}_l^{L,\alpha}(s, \tau) < \beta_l(s, \tau) < \hat{\beta}_l^{U,\alpha}(s, \tau), s \in [0, 1], \tau \in (0, 1)\} = 1 - \alpha \quad (l = 1, \dots, p).$$

By weak convergence of $\hat{\beta}(s, \tau)$, we have $\sup_{s \in [0, 1], \tau \in (0, 1)} |n^{1/2}[\hat{\beta}_l(s, \tau) - \beta_l(s, \tau) - \text{Bias}\{\hat{\beta}_l(s, \tau)\}]|$ converges weakly to $\sup_{s \in [0, 1], \tau \in (0, 1)} |\mathbb{G}_l(s, \tau)|$. De-

3.2 Simultaneous confidence bands

find $C_l(\alpha)$ such that $\text{pr}\{\sup_{s \in [0,1], \tau \in (0,1)} |\mathbb{G}_l(s, \tau)| \leq C_l(\alpha)\} = 1 - \alpha$. The $1 - \alpha$ simultaneous confidence band for $\beta_l(s, \tau)$ can be written as: $[\hat{\beta}_l(s, \tau) - \text{Bias}\{\hat{\beta}_l(s, \tau)\} - C_l(\alpha)n^{-1/2}, \hat{\beta}_l(s, \tau) - \text{Bias}\{\hat{\beta}_l(s, \tau)\} + C_l(\alpha)n^{-1/2}]$, where $\text{Bias}\{\hat{\beta}_l(s, \tau)\}$ is given by (3.5). Here, we can drop the term $\text{Bias}\{\hat{\beta}_l(s, \tau)\}$ for computational efficiency.

Now, we seek approximating $C_l(\alpha)$ by using an efficient resampling approach in line with [16], [38] and [39]. We estimate $\hat{\eta}_i(s_j) = y_i(s_j) - x_i^T \hat{\beta}(s_j, \tau)$ for $i = 1, \dots, n$ and $j = 1, \dots, m$. For $B = 1, \dots, M$, we generate independent samples $\{\varsigma_i^{(B)}, i = 1, \dots, n\}$ from $N(0, 1)$ independently and compute a stochastic process $\mathbb{G}^{(B)}(s, \tau) = -n^{-1/2}m^{-1}(I_p, 0)\check{S}_{nX}^{-1}(s, \tau) \sum_{i=1}^n \varsigma_i^{(B)} \sum_{j=1}^m K_{h_{1j}}(s) z_{x,ij}(s) \hat{w}_{ij}^*(s, \tau)$, where $\check{S}_{nX}(s, \tau) = (nm)^{-1} \sum_{i=1}^n \sum_{j=1}^m K_{h_{1j}}(s) \hat{f}_\eta(0, s_j; \tau) z_{x,ij}(s) z_{x,ij}^T(s)$ and $\hat{w}_{ij}^*(s, \tau) = I\{\hat{\eta}_i(s_j, \tau) \leq 0\} - \tau$. We calculate $\sup_{s \in [0,1], \tau \in (0,1)} |e_{l,p} \mathbb{G}^{(B)}(s, \tau)|$ for all B , and then adopt its $1 - \alpha$ empirical percentile as an estimate of $C_l(\alpha)$.

We are also interested in constructing simultaneous confidence bands for $\beta(s, \tau)$ for a given τ . In this case, we need to change the previous inference procedure by replacing $\sup_{s \in [0,1], \tau \in (0,1)} |e_{l,p} \mathbb{G}^{(B)}(s, \tau)|$ with $\sup_{s \in [0,1]} |e_{l,p} \mathbb{G}^{(B)}(s, \tau)|$.

4. Theoretical properties

In this section, we provide theoretical guarantees for the estimators and inference procedures developed in Sections 2 and 3. We first introduce some notation. For any smooth functions $f(s, \tau)$ and $g(s, t; \tau, \iota)$, let $\dot{f}(s, \tau) = df(s, \tau)/ds$, $\ddot{f}(s, \tau) = d^2f(s, \tau)/ds^2$, and $\dddot{f}(s, \tau) = d^3f(s, \tau)/ds^3$ and $g^{(a,b)}(s, t) = \partial^{a+b}g(s, t; \tau, \iota)/\partial^a s \partial^b t$, where a and b are any nonnegative integers. For a square matrix A , $\det(A)$ denotes the determinant of A . Define

$$\begin{aligned}\mu_r(K) &= \int s^r K(s) ds, \\ \mu_r [K^* \{(s-t)/h\}] &= \int u^r K(u) K\left(u + \frac{s-t}{h}\right) du, \\ \nu_r(K) &= \int s^r K^2(s) ds, \\ \mu_r(K; s, h) &= \int_0^1 h^{-r} (u-s)^r K_h(u-s) du,\end{aligned}$$

where r is any nonnegative integer.

Denote

$$\begin{aligned}\Phi(s, h) &= \begin{Bmatrix} \mu_0(K; s, h) & \mu_1(K; s, h) \\ \mu_1(K; s, h) & \mu_2(K; s, h) \end{Bmatrix}, \\ \tilde{\mu}(K, s, h_1) &= \frac{\mu_2^2(K, s, h_1) - \mu_0(K, s, h_1)\mu_3(K, s, h_1)}{\mu_0(K, s, h_1)\mu_2(K, s, h_1) - \mu_1^2(K, s, h_1)}.\end{aligned}$$

In this section, we present the conditions needed for the main theorems 1, 3 and 4. We give the following regularity conditions that are sufficient

for the following asymptotic properties, although they might not be the weakest possible. Moreover, we do not distinguish the differentiation and continuation at the boundary points from those in the interior of $[0, 1]$. For example, a function is continuous at the boundary of $[0, 1]$, which means that the function is left continuous at 0 and right continuous at 1.

The technical conditions we need are as follows.

(C1). The sequence $\{\eta_i(s, \tau) : s \in [0, 1]\}$ is a stochastic process whose τ th quantile conditional on (x_i, s) equals zero.

(C2). The grid points \mathcal{S} are prefixed according to $\pi(s)$ such that $\int_0^{s_j} \pi(s) ds = j/m$ for $1 \leq j \leq m$. Moreover, $\pi(s) > 0$ for $s \in [0, 1]$ and $\pi(s)$ has continuous second-order derivative on $(0, 1)$.

(C3). The distribution function $F_\eta(a, s; \tau)$ is Lipschitz continuous in a for each s and τ , with continuous density $f_\eta(a, s; \tau)$ uniformly bounded away from 0 and ∞ , which has continuous second-order derivative at s and is Lipschitz continuous in τ . The joint distribution function $F_\eta(a, b, s, t; \tau, \iota)$ is continuous at a and b , and has continuous second-order derivatives at s and t , with continuous density $f_\eta(a, b, s, t; \tau, \iota)$ and which has continuous second-order derivatives at s and t .

(C4). The $\beta_l(s, \tau)$ have continuous second derivatives with respect to s for each τ for all $l = 1, \dots, p$. Also, the $\partial^2 \beta_l(s, \tau) / \partial s^2$ are Lipschitz

continuous over $\{(s, \tau) : (t - s)/h_1 \in \text{supp}\{K(\cdot)\}, s, t \in \mathcal{I}, \tau \in (0, 1)\}$ for all $l = 1, \dots, p$.

(C5). The covariates x_i are independently and identically distributed with $\|x_i\| < \infty$, and $E(x_i x_i^T) = \Omega_x$ which is positive definite.

(C6). The kernel function $K(\cdot)$ in Section 2.2 of the main text is a symmetric Lipschitz continuous density function with compact support $[-1, 1]$. Moreover, $\inf_{h \in (0, h_0], s \in [0, 1]} \det\{\Phi(s, h)\} > 0$ for a small scalar $h_0 > 0$.

(C7) The bandwidth h_1 satisfies (i) $h_1 \rightarrow 0$, $mh_1 \rightarrow \infty$; (ii) $m/(nh_1^3) = O(1)$.

(C8) The density $f_\eta(a, s; \tau)$ has bounded continuous third order partial derivatives with respect to a for all $s \in [0, 1]$ and $\tau \in (0, 1)$; The kernel function $K(\cdot)$ in $K_{h_2}(\cdot)$ are third order continuously differentiable; The bandwidth h_2 satisfies $nh_2^3 \rightarrow \infty$, $nh_2^6 \rightarrow 0$ and $h_1^{-1}h_2^{2/3} \rightarrow \infty$.

(C9) The $f_\eta(a, b, s, t; \tau, \iota)$ has bounded continuous second order partial derivatives and mixed derivatives with respect to a and b for all $s, t \in [0, 1]$ and the given $\tau, \iota \in (0, 1)$; The kernel function $K(\cdot)$ in $K_{h_3}(\cdot)$ and $K_{h_4}(\cdot)$ is Lipschitz continuous; The bandwidths h_3 and h_4 satisfy $h_1 h_l^{-1} \rightarrow 0$ and $nh_l^4 \rightarrow \infty$ for $l = 3, 4$.

(C10) For the given τ and ι , the function $\tilde{\gamma}(s, t) = \gamma(s, s; \tau, \tau)^{-1/2} \gamma(s, t; \tau, \iota) \gamma(t, t; \iota, \iota)^{-1/2}$, $s, t \in [0, 1]$, has finite trace, that is, $\text{tr}(\tilde{\gamma}) = \int_0^1 \tilde{\gamma}(s, s) ds < \infty$,

where $\gamma(s, t; \tau, \iota)$ is defined in the following Theorem 1 (b).

Remark 1. Condition (C1) is a general assumption for quantile regression. Conditions (C2) and (C4)-(C7) are used by [39]. Conditions of distribution/density function are necessary for quantile regression ([14], [3], [28], [4], [13]). Condition (C3) is a very general assumption for the distribution/density function, which is specific to quantile regression for functional responses data. Conditions (C8)-(C9) are used to establish weak consistency of the error densities of $\eta(s, \tau)$, and the unknown bivariate cumulative distribution of individual effects $\eta(s, \tau)$ and $\eta(t, \iota)$, for each $s, t \in [0, 1]$ and $\tau, \iota \in (0, 1)$. For obtaining a χ^2 -type mixture distribution of T_n in (3.6), the finite trace of $\tilde{\gamma}(s, t)$ in Condition (C10) is needed, which assures that $\tilde{\gamma}(s, t)$ has the singular value decomposition as defined in (4.1). Condition (C10) is similar to the conditions used in Theorem 7 of [35].

Remark 2. Conditions (C2) and (C7) are the weak conditions on the random grid points $\mathcal{S} = \{s_j : 1, \dots, m\}$. Our proof can also be easily extended to the fixed grid points case.

We first present the global uniform Bahadur representation and weak convergence results of our estimate $\hat{\beta}(s, \tau)$ in the following theorem. The proof of this theorem is deferred to the supplementary material.

Theorem 1. *Suppose that Conditions (C1)-(C7) hold. The following results hold:*

(a) *(Global uniform Bahadur representation)*

$$\begin{aligned} & \sqrt{n} \left[\hat{\beta}(s, \tau) - \beta(s, \tau) - \frac{1}{2} h_1^2 \tilde{\mu}(K, s, h_1) \ddot{\beta}(s, \tau) \{1 + o_p(1)\} \right] \\ &= \pi(s)^{-1} f_\eta^{-1}(0, s; \tau) (I_p, 0) \{ \Phi^{-1}(s, h_1) \otimes \Omega_x^{-1} \} \\ & \quad \times \left\{ \frac{1}{\sqrt{nm}} \sum_{i=1}^n \sum_{j=1}^m [\tau - I\{\eta_i(s_j, \tau) \leq 0\}] z_{x,ij}(s) K_{h_{1j}}(s) \right\}. \end{aligned}$$

(b) *(Weak convergence)*

$$\sqrt{n} \left[\hat{\beta}(s, \tau) - \beta(s, \tau) - \frac{1}{2} h_1^2 \tilde{\mu}(K, s, h_1) \ddot{\beta}(s, \tau) \{1 + o_p(1)\} : s \in [0, 1], \tau \in (0, 1) \right]$$

converges to weakly to a two-parameter centered Gaussian process $\mathbb{G}(\cdot, \cdot)$

with covariance matrix function $\Sigma(s, t; \tau, \iota) = E\{\mathbb{G}(s, \tau)\mathbb{G}^T(t, \iota)\} = \gamma(s, t; \tau, \iota)\Omega_x^{-1}$,

where

$$\gamma(s, t; \tau, \iota) = (1, 0)A^{-1}(s, \tau)G(s, t; \tau, \iota)A^{-1}(t, \iota)(1, 0)^T,$$

$$A(s, \tau) = f_\eta(0, s; \tau)\Phi(s, h_1),$$

$$G(s, t; \tau, \iota) = \begin{pmatrix} G_{00}(s, t; \tau, \iota) & G_{01}(s, t; \tau, \iota) \\ G_{10}(s, t; \tau, \iota) & G_{11}(s, t; \tau, \iota) \end{pmatrix}$$

with

$$G_{kl}(s, t; \tau, \iota) = \begin{cases} \mu_k(K; s, h_1)\mu_l(K; s, h_1)(\tau \wedge \iota - \tau\iota), & \text{if } s = t, \\ \mu_k(K; s, h_1)\mu_l(K; t, h_1) \{F_\eta(0, 0, s, t; \tau, \iota) - \tau\iota\}, & \text{if } s \neq t, \end{cases}$$

($k, l = 0, 1$).

Remark 3. Theorem 1 (a) gives a global uniform Bahadur representation of the local quantile regression estimator $\{\hat{\beta}(s, \tau) : s \in [0, 1], \tau \in (0, 1)\}$. Theorem 1 (b) establishes the weak convergence of the $\{\hat{\beta}(s, \tau) : s \in [0, 1], \tau \in (0, 1)\}$. For Theorem 1 (b), if $s \in (0, 1)$, then $\sqrt{n} \left[\hat{\beta}(s) - \beta(s) - 0.5h_1^2\mu_2(K) \ddot{\beta}(s)\{1 + o_p(1)\} : s, \tau \in (0, 1) \right]$ converges weakly to a centered Gaussian process $\mathbb{G}(\cdot, \cdot)$ with covariance matrix $\Sigma(s, t; \tau, \iota) = \frac{F_\eta(0, 0, s, t; \tau, \iota) - \tau\iota}{f_\eta(0, s; \tau)f_\eta(0, t; \iota)}\Omega_x^{-1}$.

Theorem 2. (a) Under Conditions (C1)-(C8), we have for $s \in [0, 1], \tau \in (0, 1)$

$$\begin{aligned} & \sup_a |\hat{f}_\eta(a, s; \tau) - f_\eta(a, s; \tau)| \\ &= O_p \left\{ h_2^2 + h_1^2 + (nh_2)^{-1/2} |\log h_2|^{1/2} + n^{-1/2} + h_2^{-4} (n^{-3/2} + h_1^6) \right\}. \end{aligned}$$

(b) Under Conditions (C1)-(C8) and (C9), we have for $s, t \in [0, 1]$ and $\tau, \iota \in (0, 1)$

$$\begin{aligned} & \sup_{b_1, b_2} |\hat{f}_\eta(b_1, b_2, s, t; \tau, \iota) - f_\eta(b_1, b_2, s, t; \tau, \iota)| \\ &= O_p \left\{ h_3^2 + h_4^2 + \left(\frac{\log n}{nh_3h_4} \right)^{1/2} \right\} + O_p \left\{ (h_1^2 + n^{-1/2}) (h_3^{-2} + h_4^{-2}) \right\}. \end{aligned}$$

Theorem 2 implies that the estimators of “error” processes densities $f_\eta(\cdot, s)$ and $f_\eta(\cdot, \cdot, s, t)$ uniform converges to their true densities, and gives their rates of convergence. From Theorem 2, $\sup_{a,b \in [0,1]} |\hat{F}_\eta(a, b, s, t) - F_\eta(a, b, s, t)| = o_p(1)$ for each $s, t \in [0, 1]$, which is used in the inference procedure.

Define $\tilde{\gamma}(s, t) = \gamma(s, s; \tau, \tau)^{-1/2} \gamma(s, t; \tau, \tau) \gamma(t, t; \tau, \tau)^{-1/2}$, $s, t \in [0, 1]$. By Condition (C10), $\int_0^1 \int_0^1 \tilde{\gamma}^2(s, t) ds dt < \infty$. By Cauchy-Schwarz inequality, the function $\tilde{\gamma}(s, t)$ has the eigen decomposition

$$\tilde{\gamma}(s, t) = \sum_{l=1}^{l_0} \kappa_l \psi_l(s) \psi_l(t)$$

where the κ_l s are the eigenvalues, in decreasing order, ψ_l s are the associated orthonormal eigenfunctions of $\tilde{\gamma}(s, t)$, and l_0 is the number of positive eigenvalues.

We next present theorems about our inference procedure. The proofs of the theorems are deferred to the supplementary material.

Theorem 3. (a) *Under Conditions (C1)-(C9), we have*

$$T_n = \tilde{T}_n + o_p(1).$$

(b) *In addition, if Condition (C10) holds, then*

$$T_n^* \Rightarrow \sum_{l=1}^{l_0} \kappa_l \chi_l^2(r),$$

where $\chi_l^2(r)$ is the l th random variable which follows a central χ^2 -distribution with r degrees of freedom.

Theorem 4. Under Conditions (C1)-(C10), the bootstrapped process of $\{\mathbb{G}^{(B)}(s, \tau) : s \in [0, 1], \tau \in (0, 1)\}$ converges weakly to $\mathbb{G}(s, \tau)$ conditioning on the data, where $\mathbb{G}(s, \tau)$ is a centered Gaussian process indexed by $s \in [0, 1]$ and $\tau \in (0, 1)$.

Remark 4. Theorems 3 and 4 present the theoretical support for the statistical inference procedure introduced in Section 3. Theorem 3 suggests that T_n asymptotically follows a mixture of chi-square distribution. To use the asymptotic results for testing, we need to select a well-chosen integer ι such that the eigenvalues κ_l ($l = 1, \dots, \iota$) explain a sufficiently large portion of the total variation $tr(\tilde{\gamma})$, and this threshold is often difficult to select in practice. Therefore, we adopt approximate critical values of T_n by using the bootstrap method, where Theorem 4 provides the theoretical guarantee.

Remark 5. The proofs of the Theorems 3 and 4 rely on the uniform convergence rates of $\hat{f}_\eta(\cdot, s)$ and $\hat{f}_\eta(\cdot, \cdot, s, t)$. These results are developed in Theorem and 2.

5. Simulation study

In the section, we conduct simulation studies to evaluate our estimation and inference procedures. In the estimations and the SCBs, their finite sample performances in our KS wild bootstrap method and from LI random weighted bootstrap method of [20] (hereafter denoted by KS and LI, respectively), are compared in terms of root mean integrated squared errors, uniform coverage probabilities, average coverage widths as well as graphical visualization.

The data are generated from the following heteroscedastic model $y_i(s_j) = x_{i1}\beta_1(s_j) + x_{i2}\beta_2(s_j) + x_{i3}\beta_3(s_j) + \eta_i(s_j, \tau)$, $i = 1, \dots, n, j = 1, \dots, m$, where $x_{i1} = 1$, $(x_{i2}, x_{i3})^T \sim N\{(0, 0)^T, \text{diag}(1-1/\sqrt{2}, 1-1/\sqrt{2})+1/\sqrt{2}(1, 1)^T(1, 1)\}$, $s_j \sim \text{Uniform}[0, 1]$. To generate $\eta_i(s_j, \tau)$, we first let $\eta_i(s_j) = v_i(s_j) + \varepsilon_i(s_j)$, where $\{v_i(s_1), \dots, v_i(s_m)\}^T$ follows a multivariate normal distribution with zero mean, and its covariance matrix has a first-order autoregressive correlation structure with $\text{corr}\{v_i(s_j), v_i(s_l)\} = \gamma^{|j-l|}$ with strength $\gamma = 0.5$, $\varepsilon_i(s_j) \sim N(0, 0.2)$. As $\eta_i(s_1), \dots, \eta_i(s_m)$ are dependent, we set $\eta_i(s_j, \tau) = \eta_i(s_j) - F^{-1}(\tau)$ with F being the marginal density function of $\eta_i(s_j)$. Here $F^{-1}(\tau)$ is subtracted from $\eta_i(s_j)$ to make the τ th quantile of $\eta_i(s_j, \tau)$ zero for identifiability. We set $\beta_1(s) = s^2$, $\beta_2(s) = (1-s)^4$, and $\beta_3(s) = \exp(s) - 1$. Throughout our numerical studies, we use a Gaussian kernel and select

bandwidths using the procedure in Section 2.4.

We report the root mean integrated squared error, defined as

$$\text{RMISE}_\tau = \left\{ m^{-1} \sum_{j=1}^m |\hat{\beta}_l(s_j, \tau) - \beta_l(s_j, \tau)|^2 \right\}^{1/2} \quad (l = 1, 2, 3),$$

where m is the number of locations. The average root mean integrated squared errors over 500 Monte Carlo runs for $m = 50, 70, 90$ are included in Table 1. The results show that RMISE_τ decreases as m increases, which validate the consistency results of two estimation methods. KS outperforms LI because of having much smaller RMISE in most cases.

Second, we test the hypotheses $H_0 : \beta_1(s, \tau) = 0$ for all s against $H_1 : \beta_1(s, \tau) \neq 0$ for at least one s , for the fixed $\tau = 0.10, 0.25, 0.50, 0.75$ and 0.90 . Here, we don't compare with LI, because [20] did not give the global hypothesis test. We set $\beta_2(s) = c(1 - s)^4$ with $c = 0, 0.2, 0.3, \dots, 3.0$ to examine the power of T_n^* when $m = 50$ and $n = 100$. We use $B = 200$ bootstrap samples, and depict the power curves in Figure 1. It can be seen that the rejection rates for T_n^* based on the wild bootstrap method are accurate for the different quantile levels at both significance levels ($\alpha = 0.05$ and 0.01). But the powers of hypothesis test at the extreme quantile levels ($\tau = 0.10$ and 0.90) are obviously weaker than ones of the moderate levels ($\tau = 0.25, 0.50$ and 0.75).

Last, we evaluate the coverage probabilities (CP) and coverage widths

Table 1: Simulated results for root mean integrated squared errors for KS and LI methods.

| Method | τ | $m = 50$ | | | $m = 70$ | | | $m = 90$ | | |
|--------|--------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| | | $\beta_1(s)$ | $\beta_2(s)$ | $\beta_3(s)$ | $\beta_1(s)$ | $\beta_2(s)$ | $\beta_3(s)$ | $\beta_1(s)$ | $\beta_2(s)$ | $\beta_3(s)$ |
| KS | 0.10 | 0.0139 | 0.0295 | 0.0108 | 0.0110 | 0.0249 | 0.0094 | 0.0081 | 0.0186 | 0.0071 |
| | 0.25 | 0.0115 | 0.0291 | 0.0100 | 0.0118 | 0.0287 | 0.0077 | 0.0071 | 0.0173 | 0.0073 |
| | 0.50 | 0.0130 | 0.0279 | 0.0129 | 0.0092 | 0.0237 | 0.0115 | 0.0074 | 0.0172 | 0.0086 |
| | 0.75 | 0.0147 | 0.0294 | 0.0097 | 0.0140 | 0.0246 | 0.0131 | 0.0079 | 0.0174 | 0.0080 |
| | 0.90 | 0.0154 | 0.0308 | 0.0114 | 0.0135 | 0.0288 | 0.0096 | 0.0078 | 0.0183 | 0.0084 |
| LI | 0.10 | 0.0205 | 0.0283 | 0.0235 | 0.0187 | 0.0235 | 0.0200 | 0.0176 | 0.0225 | 0.0222 |
| | 0.25 | 0.0142 | 0.0235 | 0.0228 | 0.0144 | 0.0210 | 0.0163 | 0.0091 | 0.0167 | 0.0153 |
| | 0.50 | 0.0172 | 0.0195 | 0.0220 | 0.0142 | 0.0212 | 0.0194 | 0.0107 | 0.0156 | 0.0158 |
| | 0.75 | 0.0191 | 0.0256 | 0.0266 | 0.0151 | 0.0231 | 0.0225 | 0.0122 | 0.0169 | 0.0160 |
| | 0.90 | 0.0203 | 0.0246 | 0.0207 | 0.0264 | 0.0198 | 0.0253 | 0.0226 | 0.0215 | 0.0184 |

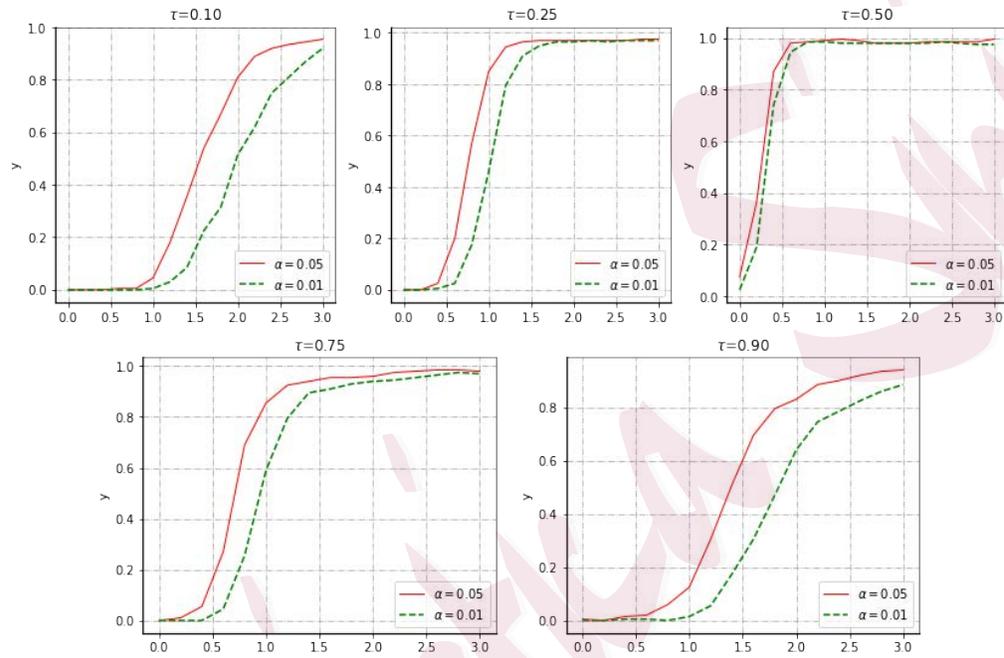


Figure 1: Plot of power curves. Rejection rate of T_n^* based on the wild bootstrap methods at $c = 0, 0.2, 0.4, \dots, 3.0$ for the different quantile levels at significance levels $\alpha = 0.05$ (real) and 0.01 (dash).

(SW) of the simultaneous confidence bands for $\beta_l(s)$ ($l = 1, 2, 3$) based on the wild bootstrap procedure, comparing with ones of LI method in [20]. Both methods have theoretical guarantee. For estimation of $C(\alpha)$, [20] adopted the weighted bootstrap method proposed in Belloni et al. [2]; We draw random weights from exponential distribution with parameter 1. In the simulation, we find the $C(\alpha)$ s obtained by [20] are slightly large, resulting in the empirical CPs significantly greater than the given nominal confidence levels and the wider SCBs. We don't know the reason. The SCBs obtained by LI method may need to be corrected because of linear interpolation. For comparison, we control the empirical CPs of LI method to the given nominal confidence levels by removing some extreme values generated by Bootstrap. We set $n = 500$ and $m = 50, 70, 90$ for $\tau = 0.10, 0.25, 0.50, 0.75$ and 0.90 , perform 500 runs, and take $B = 500$ bootstrap samples for each combination. We respectively consider confidence levels 95% and 99%, and list the simulated empirical coverage probabilities and coverage widths in Tables 2-3. From Tables 2-3, we have the following findings: (1) our KS method has much smaller RMISE compared to the LI method; (2) At all noise levels, the coverage probabilities are close to the nominal levels for both methods. (3) The coverage widths of SCBs of our KS method become narrower as the number of grid points m increases,

while the ones obtained by the LI method become much wider. That is, we have a tighter simultaneous bands when there are more sampling locations. For example, when $m = 90$, our KS method clearly outperforms the LI as our method has tighter SCBs. (4) For each approach and each regression function, the coverage widths becomes narrower as the quantile level gets closer to the median. At the extreme quantile levels $\tau = 0.10$ and 0.90 , our KS method also performs well; However, the LI method has a very wide coverage width for the given confidence level. In addition, for illustrating the good performance of our KS method, we also plot the SCBs of $\beta_i(s)$ ($l = 1, 2, 3$) for the fixed $\tau = 0.5$ at confidence levels 95% and 99% in Figure 2. The results also show that the widths of the SCBs become narrower as the number of grid points m increases. Figure 3 presents comparisons of SCBs between the KS and the LI for $\beta_i(s)$ ($l = 1, 2, 3$) at the extreme quantile level $\tau = 0.90$ and confidence level 95%. The figures for other cases are similar to them. In sum, our KS method clearly outperforms the LI method by having much smaller RMISE and much tighter simultaneous band, especially for larger sample locations.

Table 2: Simulation results for the empirical coverage probability (CP) and coverage width (CW) of 95% simultaneous confidence bands for KS and LI methods. CWs are given in parentheses.

| Method | τ | $m = 50$ | | | $m = 70$ | | | $m = 90$ | | |
|--------|--------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|
| | | $\beta_1(s)$ | $\beta_2(s)$ | $\beta_3(s)$ | $\beta_1(s)$ | $\beta_2(s)$ | $\beta_3(s)$ | $\beta_1(s)$ | $\beta_2(s)$ | $\beta_3(s)$ |
| KS | 0.10 | 0.972 (0.581) | 0.978 (0.825) | 0.962 (0.829) | 0.964 (0.493) | 0.970 (0.701) | 0.962 (0.707) | 0.990 (0.481) | 0.986 (0.681) | 0.976 (0.683) |
| | 0.25 | 0.962 (0.518) | 0.968 (0.738) | 0.944 (0.735) | 0.958 (0.439) | 0.966 (0.624) | 0.956 (0.626) | 0.972 (0.426) | 0.968 (0.605) | 0.952 (0.608) |
| | 0.50 | 0.950 (0.497) | 0.950 (0.707) | 0.932 (0.706) | 0.954 (0.421) | 0.962 (0.601) | 0.950 (0.601) | 0.968 (0.408) | 0.968 (0.580) | 0.960 (0.583) |
| | 0.75 | 0.954 (0.519) | 0.954 (0.739) | 0.948 (0.739) | 0.930 (0.437) | 0.956 (0.623) | 0.942 (0.624) | 0.972 (0.427) | 0.972 (0.605) | 0.960 (0.608) |
| | 0.90 | 0.964 (0.587) | 0.980 (0.834) | 0.970 (0.837) | 0.956 (0.494) | 0.972 (0.701) | 0.970 (0.704) | 0.986 (0.480) | 0.980 (0.681) | 0.968 (0.684) |
| LI | 0.10 | 0.928 (0.602) | 0.980 (0.863) | 0.968 (0.862) | 0.938 (0.629) | 0.972 (0.898) | 0.970 (0.900) | 0.966 (0.650) | 0.990 (0.926) | 0.986 (0.928) |
| | 0.25 | 0.926 (0.449) | 0.946 (0.654) | 0.956 (0.652) | 0.952 (0.467) | 0.966 (0.679) | 0.974 (0.681) | 0.950 (0.479) | 0.972 (0.698) | 0.974 (0.701) |
| | 0.50 | 0.944 (0.406) | 0.962 (0.594) | 0.968 (0.592) | 0.956 (0.421) | 0.970 (0.617) | 0.968 (0.616) | 0.944 (0.433) | 0.966 (0.632) | 0.968 (0.633) |
| | 0.75 | 0.968 (0.450) | 0.960 (0.654) | 0.976 (0.653) | 0.936 (0.467) | 0.958 (0.680) | 0.958 (0.680) | 0.938 (0.478) | 0.964 (0.697) | 0.956 (0.699) |
| | 0.90 | 0.942 (0.603) | 0.970 (0.864) | 0.968 (0.864) | 0.946 (0.628) | 0.984 (0.901) | 0.974 (0.901) | 0.970 (0.649) | 0.982 (0.929) | 0.982 (0.930) |

Table 3: Simulation results for the empirical coverage probability (CP) and coverage width (CW) of 99% simultaneous confidence bands for KS and LI methods. CWs are given in parentheses.

| Method | τ | $m = 50$ | | | $m = 70$ | | | $m = 90$ | | |
|--------|--------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|
| | | $\beta_1(s)$ | $\beta_2(s)$ | $\beta_3(s)$ | $\beta_1(s)$ | $\beta_2(s)$ | $\beta_3(s)$ | $\beta_1(s)$ | $\beta_2(s)$ | $\beta_3(s)$ |
| KS | 0.10 | 0.994 (0.738) | 0.996 (1.049) | 0.990 (1.056) | 0.992 (0.622) | 0.992 (0.887) | 1.000 (0.892) | 0.998 (0.608) | 0.992 (0.863) | 0.992 (0.864) |
| | 0.25 | 0.986 (0.658) | 0.988 (0.937) | 0.978 (0.936) | 0.992 (0.554) | 0.990 (0.792) | 0.988 (0.794) | 0.990 (0.539) | 0.988 (0.767) | 0.996 (0.772) |
| | 0.50 | 0.986 (0.629) | 0.986 (0.898) | 0.978 (0.900) | 0.988 (0.533) | 0.986 (0.758) | 0.988 (0.761) | 0.996 (0.519) | 0.990 (0.735) | 0.992 (0.738) |
| | 0.75 | 0.992 (0.659) | 0.988 (0.942) | 0.984 (0.943) | 0.988 (0.555) | 0.988 (0.788) | 0.992 (0.789) | 0.992 (0.540) | 0.988 (0.763) | 0.992 (0.769) |
| | 0.90 | 0.992 (0.743) | 0.992 (1.056) | 0.994 (1.063) | 0.986 (0.626) | 0.994 (0.889) | 0.994 (0.891) | 0.994 (0.607) | 0.986 (0.861) | 0.990 (0.866) |
| LI | 0.10 | 0.984 (0.680) | 0.992 (0.966) | 0.990 (0.964) | 0.990 (0.708) | 0.996 (1.000) | 0.996 (1.001) | 0.996 (0.771) | 1.000 (1.080) | 0.998 (1.082) |
| | 0.25 | 0.972 (0.497) | 0.992 (0.726) | 0.986 (0.723) | 0.986 (0.514) | 0.992 (0.749) | 0.992 (0.751) | 0.996 (0.551) | 1.000 (0.803) | 0.998 (0.807) |
| | 0.50 | 0.980 (0.448) | 0.984 (0.658) | 0.998 (0.656) | 0.988 (0.462) | 0.992 (0.677) | 0.990 (0.678) | 0.994 (0.494) | 0.992 (0.725) | 0.996 (0.726) |
| | 0.75 | 0.988 (0.498) | 0.990 (0.726) | 0.996 (0.724) | 0.978 (0.514) | 0.992 (0.750) | 0.992 (0.751) | 0.994 (0.551) | 0.996 (0.802) | 0.996 (0.805) |
| | 0.90 | 0.984 (0.681) | 0.998 (0.968) | 0.998 (0.969) | 0.990 (0.707) | 0.998 (1.004) | 0.988 (1.004) | 0.996 (0.771) | 0.998 (1.083) | 1.000 (1.086) |

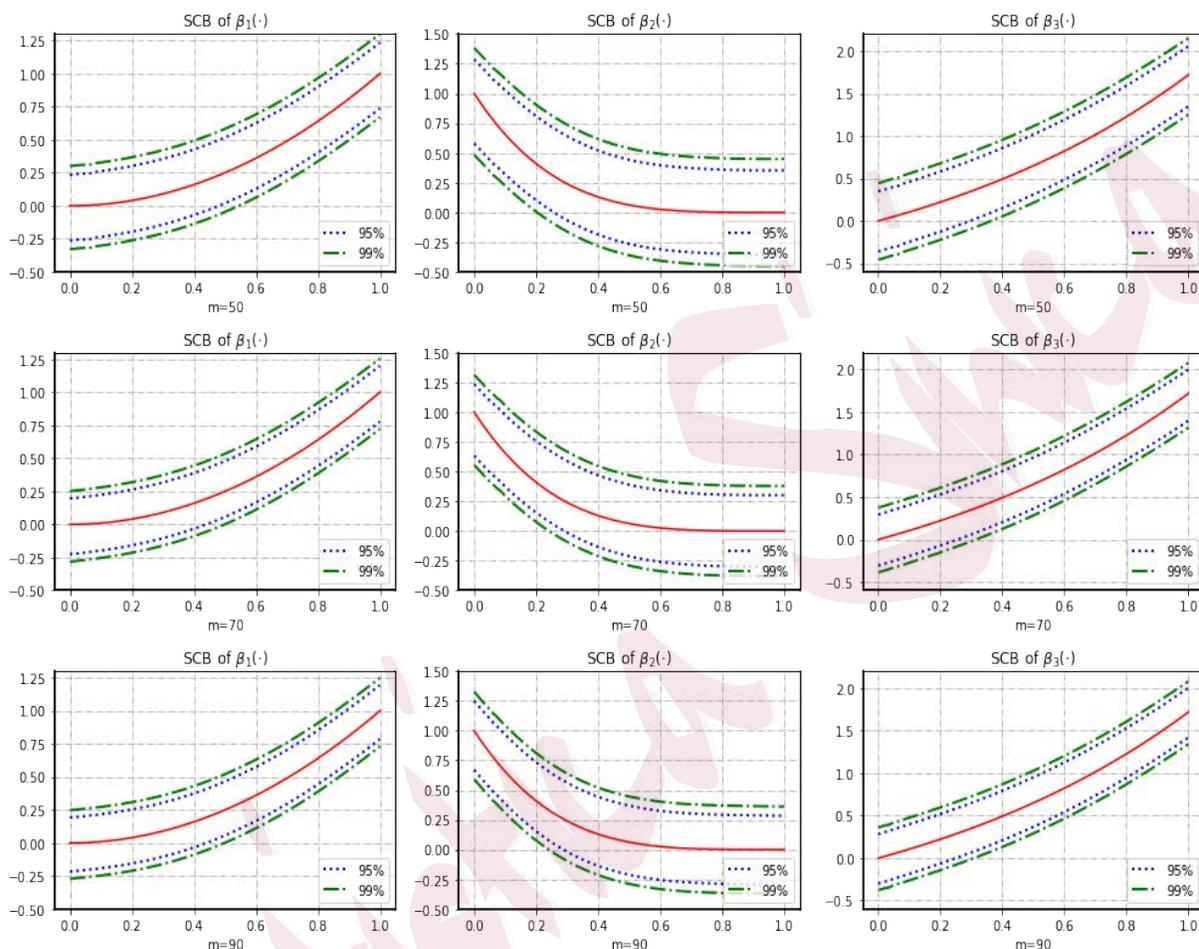


Figure 2: The simultaneous confidence bands, denoted by “SCB”, of three coefficient functions for the $\tau = 0.5$ when nominal levels are 95% and 99%, based on the KS method. The red dotted curves are the true coefficient functions, the blue dot curves and the green dash-dot curves respectively are the confidence bands of nominal levels 95% and 99%.

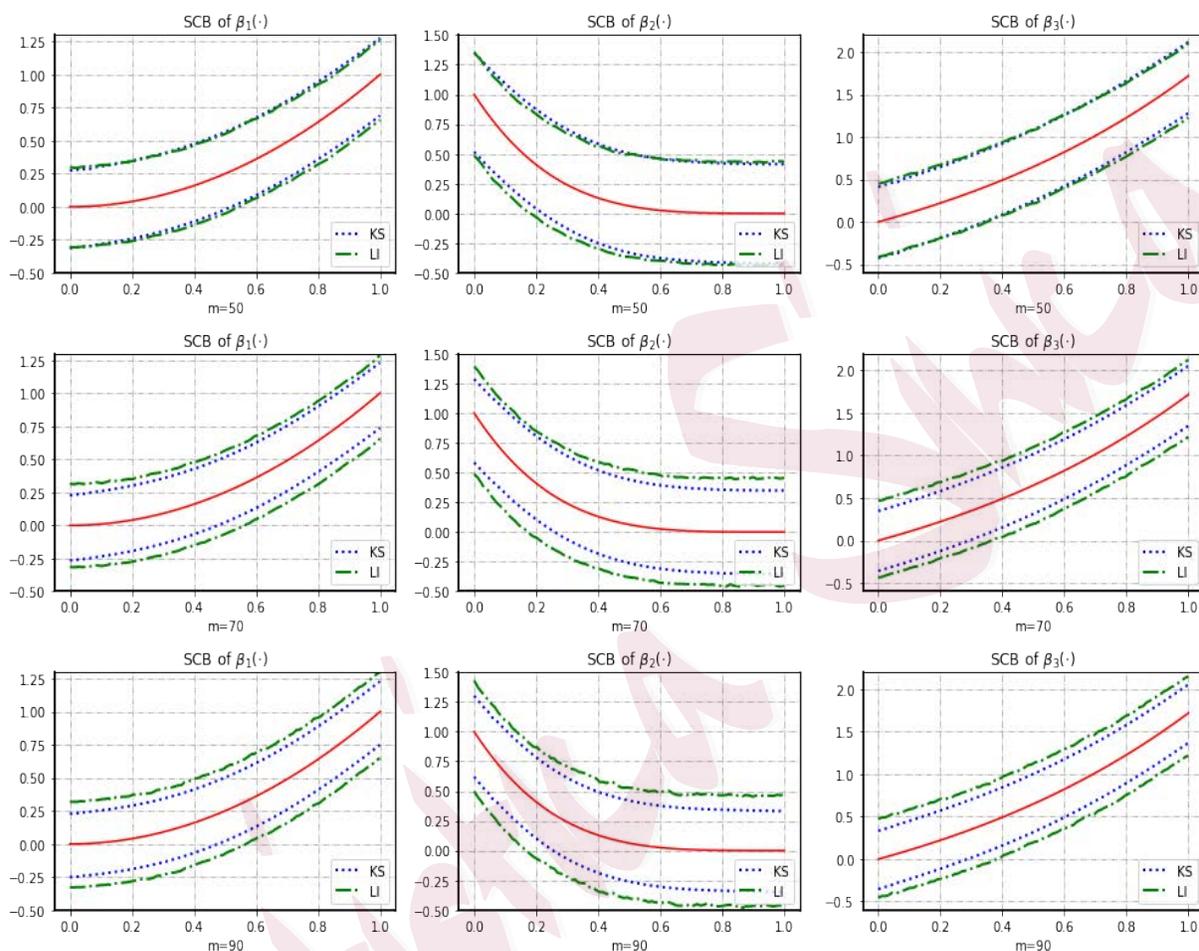


Figure 3: Comparisons of SCBs between the KS and the LI for three coefficient functions at quantile level $\tau = 0.90$ and nominal level 95%. The red dotted curves are the true coefficient functions, the blue dot curves and the green dash-dot curves respectively are the confidence bands of the KS and the LI.

6. Diffusion tensor imaging data analysis

We analyze imaging data from the neonatal project on early brain development. The data set consists of 128 healthy infants, including 75 males and 53 females. The gestational ages of the infants range from 262 to 433 days. The diffusion tensor imaging and T1-weighted images were acquired for each subject. We have included the detailed imaging preprocessing steps in the supplementary material.

We consider two diffusion properties: fractional anisotropy and mean diffusivity, which were measured at 45 grid points along the genu tract of the corpus callosum. They measure the inhomogeneous extent of local barriers to water diffusion and the averaged magnitude of local water diffusion, respectively. At different quantile levels τ , we are interested in exploring the effects of the infant's gender and the gestational age at magnetic resonance imaging scanning on the fractional anisotropy and mean diffusivity, and to delineate the tendency of fiber diffusion properties across time. For $\tau = 0.25, 0.5, 0.75$, we respectively fit our model by taking fractional anisotropy and mean diffusivity values as response, and gender and age as covariates. We also include an intercept term. The age, fractional anisotropy and mean diffusivity values are standardized before fitting the model (1.2). For each given τ , we estimate the coefficient function $\beta(s) = \{\beta_0(s), \beta_1(s), \beta_2(s)\}^T$,

compute T_n^* for each hypothesis test and obtain the p -value by applying the wild bootstrap procedures with $B = 500$ replications.

For the given quantile levels, Figure 4 plots the estimated coefficient functions corresponding to intercept, gender and age associated with fractional anisotropy [panels (a) (b) (c)] and mean diffusivity [panels (d), (e), (f)], and p -values of global test statistics under null hypothesis $H_{0l} : \beta_l(s) = 0$ ($l = 1, 2, 3$). The intercept functions $\beta_0(s)$ [panels (a) and (d)] give the overall tendency, and are significantly grid-point-varying at all quantiles considered at the 0.01 level. In addition, from panels (d), we see that the intercept function corresponding to the mean diffusivity in lower quantile $\tau = 0.25$, that is, for the group of healthy infants with the lower mean diffusivity, is negative, whereas that corresponding to the group of healthy infants with higher mean diffusivity is positive. Such negative and positive effects would not have been revealed by methods focusing only on the conditional mean. For gender effects [panels (b) and (e)], the six p -values are all greater than 0.01 except for mean diffusivity when $\tau = 0.5$. This is consistent with findings in [39]; gender has a weakly significant effect on fractional anisotropy and mean diffusivity. Male infants only have relatively bigger averaged magnitudes of local water diffusion along genu tract of the corpus callosum compared with female infants for first 5 grid points

when $\tau = 0.5$. For gestational age effects [panels (c) and (f)], the p -value is greater than 0.01 for the fractional anisotropy; while it is less than 0.01 for the mean diffusivity response. This indicates that the gestational age has a weakly significant effect on the fractional anisotropy along the genu tract of the corpus callosum; while the mean diffusivity along the genu tract of the corpus callosum changes noticeably with the gestational age.

7. Discussion

In the paper, we have proposed a functional response quantile regression model, which explicitly characterizes the conditional distribution of functional response given scalar predictors. We have developed a global test statistic for linear hypotheses of the varying coefficient functions, and constructed an asymptotic SCB for each regression coefficient function. Simulations and real data analysis show that the FRQR can characterize the effect of scalar predictors on the functional responses at different quantile levels.

We focus on densely observed functional data in this paper. And in practice, the functional data may be irregularly and sparsely observed. Our method is not directly applicable to this case because the strong approximation results used to construct nonparametric confidence bands, commonly

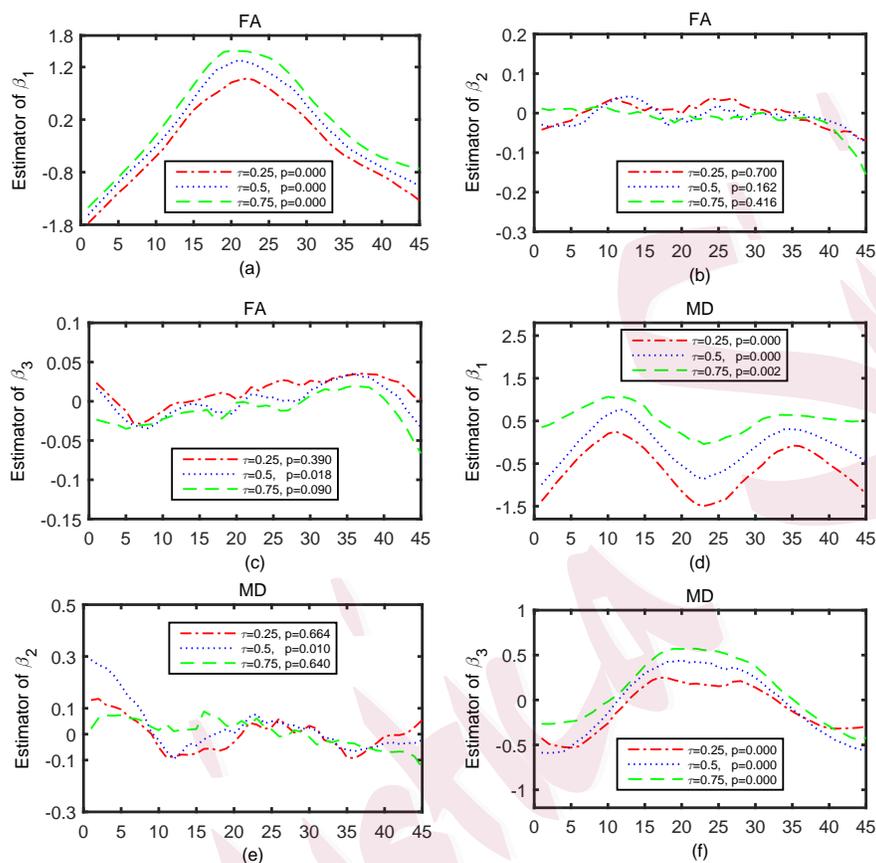


Figure 4: Plot of estimated coefficient functions and p -values of global test statistics for the quantile levels $\tau = 0.25, 0.5$ and 0.75 for diffusion tensor imaging data. “FA” denotes fractional anisotropy and “MD” denotes mean diffusivity.

REFERENCES

known as “Hungarian embedding”, can not be applied to the irregular and sparse functional data. As far as we are aware, only pointwise confidence bands have been developed under the irregular and sparse functional data context [32, 33, 31, 21, 37, 5]. It would be an interesting direction to explore how to construct SCBs for such data in the future. In addition, it is also an important issue to develop a theory to support the procedure with tuning parameters selected adaptively; The issue is difficult and remains open.

Supplementary Material

The supplementary material includes all of proofs and technical details.

Acknowledgements

Dr. Zhou was supported by the National Social Science Fund of China (19BTJ034) and the National Natural Science Foundation of China (12171242, 11971235). We are grateful to the editor, associate editor, and the anonymous referees for their many helpful comments and suggestions.

References

- [1] M.G. Akritas and I. Van. Keilegom. Nonparametric estimation of the residual distribution. *Scand. J. Statist.*, 28:549–567, 2001.

REFERENCES

- [2] A. Belloni, V. Chernozhukov, D. Chetverikov, and I. Fernández-Val. Conditional quantile processes based on series or many regression. *Journal of Econometrics*, 213(1):4–29, 2019.
- [3] Z. Cai and X. Xu. Nonparametric quantile estimations for dynamic smooth coefficient models. *J. Am. Statist. Assoc.*, 104:371–383, 2009.
- [4] K. Chen and H.G. Müller. Conditional quantile analysis when covariates are functions with application to growth data. *J. R. Statist. Soc. B*, 74:1–23, 2012.
- [5] D. Şentürk and H.-G. Müller. Functional varying coefficient models for longitudinal data. *Journal of the American Statistical Association*, 105:1256–1264, 2010.
- [6] J. Fan and I. Gijbels. *Local Polynomial Modelling and Its Applications*. Chapman & Hall, London, 1996.
- [7] J. Fan, Q. Yao, and H. Tong. Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems. *Biometrika*, 83:189–206, 1996.
- [8] J. Fan and T.H. Yim. A data-driven method for estimating conditional densities. *Biometrika*, 91:819–834, 2004.
- [9] Marta I Garrido, Maneesh Sahani, and Raymond J Dolan. Outlier responses reflect sensitivity to statistical structure in the human brain. *PLoS Comp. Biology*, 9(3):e1002999, 2013.
- [10] P. Hall, J. Racine, and Q. Li. Cross-validation and the estimation of conditional probability densities. *J. Am. Statist. Assoc.*, 99:1015–1026, 2004.

REFERENCES

- [11] T. Hastie and R. Tibshirani. Varying-coefficient models. *J. R. Statist. Soc. B*, 55:757–796, 1993.
- [12] W. Hendricks and R. Koenker. Hierarchical spline models for conditional quantiles and the demand for electricity. *J. Am. Statist. Assoc.*, 87:58–68, 1992.
- [13] K. Kato. Estimation in functional linear quantile regression. *Ann. Statist.*, 40:3108–3136, 2012.
- [14] R. Koenker. *Quantile Regression*. Cambridge Univ. Press, Cambridge, 2005.
- [15] R. Koenker and G. J. R. Bassett. Regression quantiles. *Econometrica*, 46:33–50, 1978.
- [16] M. R. Kosorok. Bootstraps of sums of independent but not identically distributed stochastic processes. *J. Mult. Anal.*, 84:299–318, 2003.
- [17] M.R. Kosorok. *Introduction to Empirical Processes and Semiparametric Inference*. Springer, New York, 2008.
- [18] Matthias Krauledat, Guido Dornhege, Benjamin Blankertz, and Klaus-Robert Müller. Robustifying eeg data analysis by removing outliers. *Chaos and Compl. Letters*, 2(3):259–274, 2007.
- [19] Y.S. Liu, M. Li, and J.S. Morris. Function-on-scalar quantile regression with application to mass spectrometry proteomics data. *Annal of Applied Statistics*, 14:521–541, 2020.
- [20] Y.S. Liu, M. Li, and J.S. Morris. On function-on-scalar quantile regression. *arXiv*, 2002.03355v1, 2020.

REFERENCES

- [21] S. Ma, L. Yang, and R. J. Carroll. A simultaneous confidence band for sparse longitudinal regression. *Statist. Sinica*, 22:95–122, 2012.
- [22] U. U. Müller, A. Schick, and W. Wefelmeyer. Estimating the error distribution function in semiparametric regression. *Statist. Decisions*, 25:1–18, 2007.
- [23] N. Neumeier and I. V. Keilegom. Estimating the error distribution in nonparametric multiple regression with applications to model testing. *J. Mult. Anal.*, 101:1067–1078, 2010.
- [24] J. O. Ramsay and C. J. Dalzell. Some tools for functional data analysis (with discussion). *J. R. Statist. Soc. B*, 53:539–572, 1991.
- [25] J. O. Ramsay and B. W. Silverman. *Functional Data Analysis (2nd ed)*. Springer, New York, 2005.
- [26] J. A. Rice and B. W. Silverman. Estimating the mean and covariance structure nonparametrically when the data are curves. *J. R. Statist. Soc. B*, 53:233–243, 1991.
- [27] Q. Shen and J. Faraway. An f test for linear models with functional responses. *Statist. Sinica*, 14:1239–1257, 2004.
- [28] H. Wang, Z. Zhu, and J. Zhou. Quantile regression in partially linear varying coefficient models. *Ann. Statist.*, 37:3841–3866, 2009.
- [29] Jane-Ling Wang, Jeng-Min Chiou, and Hans-Georg Müller. Functional data analysis. *Annual Review of Statistics and Its Application*, 3:257–295, 2016.

REFERENCES

- [30] H. Yang, V. Baladandayuthapani, A.U.K. Rao, and J.S. Morris. Quantile function on scalar regression analysis for distributional data. *Journal of the American Statistical Association*, 115:90–106, 2020.
- [31] F. Yao. Asymptotic distributions of nonparametric regression estimators for longitudinal or functional data. *J. Mult. Anal.*, 98:40–56, 2007.
- [32] F. Yao, H. G. Müller, and J. L. Wang. Functional data analysis for sparse longitudinal data. *J. Am. Statist. Assoc.*, 100:577–590, 2005.
- [33] F. Yao, H. G. Müller, and J. L. Wang. Functional linear regression analysis for longitudinal data. *Ann. Statist.*, 33:2873–2903, 2005.
- [34] J. T. Zhang. Statistical inference for linear models with functional responses. *Statist. Sinica*, 21:1431–1451, 2011.
- [35] J. T. Zhang and J. Chen. Statistical inferences for functional data. *Ann. Statist.*, 35:1052–1079, 2007.
- [36] Z.W. Zhang, X. Wang, L.L. Kong, and H.T. Zhu. High-dimensional spatial quantile function-on-scalar regression. *Journal of the American Statistical Association*, doi.org/10.1080/01621459.2020.1870984:1–16, 2021.
- [37] S. Zheng, L. Yang, and W. Härdle. A smooth simultaneous confidence corridor for the mean of sparse functional data. *Journal of the American Statistical Association*, 109:661–673, 2014.

REFERENCES

[38] H. T. Zhu, J. G. Ibrahmt, N. Tang, D. B. Rowe, X. Hao, R. Bansal, and B. S. Peterson.

A statistical analysis of brain morphology using wild bootstrapping. *IEEE Trans. Med. Imaging*, 26:954–966, 2007.

[39] H. T. Zhu, R.Z. Li, and L.L. Kong. Multivariate varying coefficient model for functional

responses. *Ann. Statist.*, 40:2634–2666, 2012.

Xingcai Zhou, Institute of Statistics and Data Science, Nanjing Audit University, Nanjing 211815, China

E-mail: xczhou@nuaa.edu.cn

Dehan Kong, Department of Statistical Sciences, University of Toronto, Toronto, Ontario M5S 3G3, Canada

E-mail: kongdehan@utstat.toronto.edu

A. B. Kashlak, Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta T6G 2G1, Canada

E-mail: kashlak@ualberta.ca

Linglong Kong, Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta T6G 2G1, Canada

E-mail: lkong@ualberta.ca

R. Karunamuni, Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta T6G 2G1, Canada

E-mail: r.j.karunamuni@ualberta.ca

Hongtu Zhu, Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27516, U.S.A.

E-mail: hzhu@bios.unc.edu